Consiglio Nazionale delle Ricerche

Social-Aware Stateless Forwarding in Pocket Switched Networks

A. Mei, G. Morabito, P. Santi, J. Stefa

IIT TR-22/2010

Technical report

Agosto 2010

Istituto di Informatica e Telematica

Social-Aware Stateless Forwarding in Pocket Switched Networks

Alessandro Mei, Giacomo Morabito, Paolo Santi, and Julinda Stefa

***Abstract*—Several social-aware routing protocols for pocket switched networks have been recently introduced in the literature. The main idea underlying these protocols is to exploit state information (e.g., history of past encounters) to deduce information on the social structure of the network, and to optimize routing based on this information. While social-aware routing protocols have been shown to have superior performance to social-oblivious, stateless routing protocols such as, e.g., BinarySW, the improvement comes at the cost of considerable storage overhead required on the nodes, which is instead not required for stateless approaches. So, whether the benefits of social-aware routing protocols would still be present when storage capacity at the nodes is constrained is not clear.**

In this paper we present SANE, the first forwarding mechanism that combines the advantages of both *social-aware* and *stateless* approaches. SANE is based on the observation—that we validate on real-world traces—that individuals with similar interests tend to meet more often. In our approach, individuals (network members) are characterized by their *interest profile*, a compact representation of their interests. By implementing a simple interest profile similarity based forwarding rule, SANE is free of network state information, thus overcoming the storage capacity problem with existing social-aware approaches. Through extensive experiments, we show the superiority of social-aware, stateless forwarding over existing stateful, social-aware and stateless, social-oblivious routing approaches. An important byproduct of our interest-based approach is that it easily enables innovative routing primitives, such as interest-casting. An interest-casting protocol is also introduced in this paper, and extensively evaluated through experiments based on both real-world and synthetic mobility traces.

I. Introduction

The vision of a near future in which a multitude of hand-held devices establish direct wireless communication links in an opportunistic fashion has recently attracted the attention of the research community. This vision is motivated by the fact that powerful hand-held devices are becoming increasingly popular (smart phones, PDAs, etc.), and that these devices are typically endowed also with wireless technologies allowing direct communication between them (e.g., Bluetooth).

The above vision has motivated researchers to focus on a specific type of delay tolerant network, called *opportunistic* or *pocket switched network* [13], in which nodes are *individuals* carrying such powerful hand-held devices. Given that node mobility, coupled with a *store-carry*-and-*forward* mechanism on the nodes, is the fundamental mean of communication in delay tolerant networks in general, and in PSNs in particular, several authors have tried to exploit the fact that mobile nodes are indeed individuals characterized by social relationships to optimize communication within the network. The characterization of social ties between nodes has been used to optimize performance of unicast communications [5], [11], [16], as well as multicasting [8], and publish-subscribe mechanisms [1], [4], [14].

While social-aware routing protocols have been shown to have superior performance to social-oblivious routing protocols such as, e.g., BinarySW [21], this performance improvement comes at the expense of storing a significant amount of state information (e.g., history of past encounters, portion of the “social network” graph, etc.) at the local memory of the nodes. In other words, a common feature of the social-aware routing approaches introduced so far is that they heavily build upon a notion of *state*.

Given that existing routing approaches for PSNs have both pros and cons, it would be interesting to design a routing approach that combines the advantages of both approaches, while reducing the cons as much as possible. In particular, our goal in this paper is to design a *social-aware*, *stateless* routing approach, which combines the advantages of social-aware forwarding with the negligible extra storage requirements typical of a stateless approach. To the best of our knowledge, the one presented in this paper is the first routing approach with these features presented in the opportunistic networking literature.

Our routing approach is based on a simple forwarding mechanism, which we call SANE (Social-Aware NEtworking), exploiting the observation, qualitatively well-known in sociology [17], that individuals with similar interests tend to meet more often. This observation has been recently indirectly validated in [20], where the authors show that mobility patterns can be used to accurately predict individual interests. Indeed, a first significant contribution of this paper is a *quantitative* and *direct* validation of this observation, based on the only real-world mobility trace enriched with user profiles information we are aware of [11], [12].

We start by giving a model for representing user interests and their similarity. In our approach, the collection of interest are represented in an m-dimensional *interest space*, and the individuals of the network are characterized by their *interest profile*, an m-dimensional vector corresponding to a point in the interest space. The forwarding strategy is then driven by a measure of similarity between interest profiles which, at least indirectly, expresses strength of social ties between the corresponding individuals. Through extensive experiments based on both real-world and synthetic mobility traces, we show the superiority of our proposed social-aware, stateless routing approach over existing stateful, social-aware as well as stateless, social-oblivious routing approaches.

An important byproduct of our interest-based approach to routing is that it easily allows realizing *innovative* routing primitives, such as interest-casting. In interest-casting, a message M circulating in the network is characterized by a *message relevance profile*, represented as a point in the interest space as well, and the goal is to deliver a copy of M to all potentially interested users, i.e., those individuals whose interest profile is "close enough" (according to a certain similarity metric) to M's relevance profile. An interest-casting protocol is also introduced in this paper, and extensively evaluated through experiments based on both real-world and synthetic mobility traces.

II. Related work and contribution

The idea of exploiting information regarding social ties between network nodes in PSNs is not new. For instance, in [5] the authors use the notions of "ego-centric betweenness" and "social similarity" to improve end-to-end routing performance. In [11], the authors propose to use a social "centrality" metric to achieve the same purpose. In [16], the authors use a "social similarity" metric locally computed from the history of past encounters to route messages within the network. Recently, a social-based approach based on a notion of "ego-centric betwenness" has been proposed also to optimize multicast performance [8].

The above protocols have shown how the social structure of a PSN can be successfully exploited to improve traditional, social oblivious approaches. However, existing social-aware approaches heavily build upon the ability of storing a large amount of information at the nodes (typically, to keep trace of past encounters), i.e., their are *stateful* approaches. This fact has important implications for what concerns $i)$ scalability and $ii)$ effects of memory size on routing performance. As for $i)$, we observe that relying on a rich state (in some cases, $O(n^2)$ storage capacity is required at the nodes, where n is the number of network nodes) might impose severe limits to the ability of these approaches to scale up to networks of even medium size. As for $ii)$, to the best of our knowledge, the effect of limited memory size on social-aware routing performance has not been investigated so far. Considering limited memory size when comparing performance of stateful approaches (such as, e.g., [5], [11], [16]) to that of stateless, social-oblivious approaches such as Epidemic [22] and Binary SW [21] is very important for the following reasons. First, using memory to store state information (even if suitably compacted by, e.g., using meta-data) clearly reduces the amount of memory that can be used to store the messages circulating in the network, with a negative effect on routing performance. Second, current social-aware approaches tend to convey messages towards relatively few "socially well-connected" nodes, which could then become hotspot and incur serious buffering problems. Hence, comparing routing performance without taking the effect of limited memory size into account gives an unfair advantage to stateful approaches over stateless one. Given $i)$ and $ii)$, whether social-aware approaches are actually effective in improving routing performance is still not clear, as well as their scalability properties.

In order to at least partially address the above issues with current social-aware routing approaches, in this paper we advocate a different perspective on how information related to the user social behavior is used to optimize PSN routing performance. In particular, we propose to characterize each individual belonging to the network with an *interest profile* belonging to the network's *interest space* (see Section III for formal definitions), and to base the forwarding strategy of the routing protocol upon a similarity metric between individual interest profiles: when individual A carrying a message M destined to individual D meets another individual B, he/she compare D and B interest profiles, and, based on the outcome of this comparison, he/she decides whether to forward M to B. It is important to observe that this forwarding approach is *stateless*, since A discards B's profile after the forwarding decision has been taken. . By stateless, we mean that the amount of information used by nodes to forward messages within the network is limited to knowledge of the destination's interest profile (address), and the own interest profile. This is sharp contrast with existing social-aware routing protocols which, beyond knowing the destination's address, require storing $O(n)$ or even $O(n^2)$ extra information at each node.

The above described interest-based forwarding mechanism not only addresses issues with current social-aware unicast routing approaches, but can also be easily extended to realize *novel* networking paradigms for PSNs, which naturally build on top of the notion of interest profile and interest space. In this paper, we present one such novel paradigm, namely *interest-casting*. In interest-casting, a message M is characterized by a *relevance profile* describing its relevance to the various topics of interest/communities present in the network, and the message is destined to all network members whose interest profile "matches" M's relevance profile.

Summarizing, the contributions of this paper are:

- a first quantitative assessment, based on real-world mobility traces, of the degree of correlation between individual meeting rates and similarity of their interests;
- the design of the first social-aware, stateless forwarding approach for PSNs;
- the introduction of a novel networking primitive (interest-cast) for PSNs based on our proposed forwarding approach;
- a thorough performance assessment (based on both real-world and synthetic mobility traces) of the proposed unicast and interest-cast approaches against both stateless, social-oblivious approaches and stateful, social-aware approaches.

The line of research closer to the ideas presented in this paper is the design of social-based publish-subscribe mechanisms for PSNs, such as the ones presented in [1], [14], [4]. Among these works, the one that is most closely related to ours is [4], in which the authors present a routing mechanism, called SocialCast, that exploits predictions based on metrics of social interactions to drive the forwarding process. While the

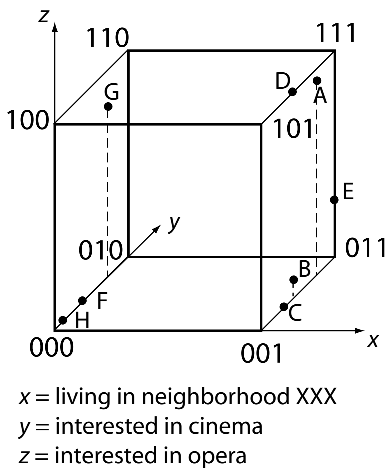


Fig. 1. An example of network with 3 interest dimensions and 8 individuals.

underlying idea of interest-based routing is similar in spirit to our approach, implicit in SocialCast is the assumption that an individual implicitly or explicitly subscribes to one or more "interests". On the contrary, our approach builds on the notion of interest profile, in which a PSN member compactly encodes not only the *degree* (not necessarily binary) of interest in different topics, but also his/her habits (e.g., where he/she lives, works, etc.), etc. Thus, our approach allows a more complete characterization of a PSN member's habits and social relationships. Finally, SocialCast still remains a publish-subscribe scheme and requires storing of a considerable amount of state information at the nodes, which should be contrasted with the stateless approach taken herein.

III. Interest space and profiles

A. The model

We assume each individual in the network can be represented through his/her *interest profile*, i.e., a compact representation of his/her interests within the *interest space*. We represent the interest space as an m-dimensional unit cube $\mathcal{C} = [0,1]^m$, where m is the total number of interests in the network under consideration. Interests are intended in a very broad sense, they might represent degree of interest in a certain topic (e.g., cinema, literature, etc.), the fact that an individual belongs to a certain physical or virtual community (e.g., living in a certain neighborhood, member of a Facebook interest group, etc.), and so on. Note that, for a given $i \in \{1,\ldots,m\}$, the value of the corresponding dimension in the interest space – i.e., the i-th interest dimension – can be either a 0/1 value or an arbitrary real value in the $[0,1]$ interval. This is to enable 0/1 interests such as "membership to a certain community", as well as arbitrary "degree of interest" in a certain topic.

Given the above definition of interest space, it is quite natural to represent the interest profile of an individual A with an m-dimensional vector reporting, for each possible interest dimension, A's degree of interest in the particular topic/community (either a real number or a binary value). Thus, we can think of individual interest profiles as points in the m-dimensional interest space. For example, Figure 1 represents a set of 8 individuals, denoted as A, B, ..., H, in a network with 3 interest dimensions. One of the communities – "living in neighborhood X" – allows only binary membership values and is represented along the x-axis, whereas the other two – "interested in cinema" and "interested in opera" – have continuous membership values and are represented along the y and z-axis, respectively. In this example, individuals A,...,E live in neighborhood X, and have different degrees of interest in cinema and opera, while individuals F, G, and H live outside neighborhood X.

To express similarity between individual interests, and thus quantitatively measure "homophily" – degree of interest similarity [17], we use the well-known *cosine similarity* metric [6]:

Definition 1: Given two m-dimensional vectors A and B, the cosine similarity metric, denoted $\Theta(A,B)$, is defined as follows:

$$\Theta(A,B) = \cos(\angle AB) = \frac{A \cdot B}{\parallel A \parallel \parallel B \parallel} ,$$

where $\parallel X \parallel$ represent the length of vector X.

Note that, given the definition of interest space, $0 \leq \Theta(A,B) \leq 1$ in our model, with higher values of $\Theta(A,B)$ corresponding to a higher "homophily" degree.

B. Validation

As described in the previous section, our stateless protocols are based on a simple and natural observation from everyday life: Our movements are guided in a large part by our interests. To validate this intuition in a quantitative fashion, we use traces collected during an experiment done with real Bluetooth communicating devices distributed to part of the participants of the Infocom 2006 conference [11], [12]. This data trace contains not only contact logs, but it also reports information on participants' nationality, residence, languages spoken, affiliation, scientific interests, etc.. From this information we can easily generate an interests profile vector of 0/1 coordinates: We count all the possible nationalities, countries and cities of residence, languages spoken, affiliations, possible scientific interest topics, declared by the participants. Then, we build, for each participant, a profile vector that has as many coordinates as the sum of all these possibilities put together. A 1 in the i-th coordinate of a given participant's profile vector corresponds to the fact that that participant is either interested in the scientific topic, or speaks that particular language, or comes from that particular country (depending on what interest dimension i represents). In the process, we discard participants that have not declared any of the above interests, in order to remove erroneous profiles. The number of the participants involved after this cut reduces to 61. Although there are other data-traces available on line describing contact among participants in different experimental settings ([7], [11], [12], [13]), they do not include any information on participants' profiles. To the best of our knowledge, Infocom 06 is the only available data-trace that includes also these type of information, thus in this paper we focus on this data-trace. More details about the data-trace can be found in Table I.

Experimental data set	Infocom 06
Device	iMote (Bluetooth)
Duration (days)	3
Granularity (sec)	120
Participants number	78
Participants with profile	61

TABLE I
DETAILED INFORMATION ON THE INFOCOM 06 TRACE.

AVG meet time	C_d	C_f	Nodes
> 0 (min)	.28	.08	61
> 5 (min)	.55	.57	53
> 10 (min)	.67	.67	26

TABLE II
CORRELATION BETWEEN INTERESTS PROFILES AND PARTICIPANTS' ENCOUNTERS. C_d AND C_f INDICATE THE PEARSON CORRELATION COEFFICIENT BETWEEN PARTICIPANTS' COUPLES PROFILES AND RESPECTIVELY TOTAL MEETING DURATIONS AND MEETING RATES.

To support our intuition, we first calculate the cosine similarity between the interest profiles for every pair of participants. Then, we compute the Pearson correlation index among this value and the total meeting duration/meeting frequency among every couple. These values result to be .28 and .08, respectively. The second correlation coefficient is small: This is more than reasonable, being this trace the result of the mobility pattern in a big conference, where there is a high "mixing" of people and thus a high number of short-casual meetings, for example, almost all the attendees meet during the coffee break. Yet, the first correlation coefficient (the one related to the duration of the contacts between people) shows that even in the presence of a high number of casual meetings, people with similar profile tend to meet for longer times. To confirm this observation, we then compute the correlation coefficients among profile similarities and meeting duration/meeting frequency, only for pairs of individuals who spend, on the average, more than a certain amount of time together. This way the effect of the casual short meetings is attenuated. The results are presented in Table II. As can be seen, when we focus on longer meetings, the correlation of meeting frequency and similarity of interest profiles is considerably high, reaching 0.67. These results support the conclusion that our intuition is sound and that it can be used as the basic mechanism of social-aware, stateless forwarding protocols.

IV. SOCIAL AWARE NETWORKING (SANE)

In this section, we introduce *Social Aware NEtworking* (SANE), a protocol suite that enables the efficient delivery of information to relevant destinations in PSNs. SANE supports a novel communication service, that we call *interest-cast* (see Section IV-B), besides the traditional unicast.

We assume that each node can be a forwarder and therefore, according to the *store-carry*-and-*forward* discipline, maintains a buffer of messages that must be relayed to the respective destinations. Each message M has a header that contains a target interest profile that we call *message relevance profile*, an integer value N_{replicas} representing the number of replicas of the message that the node is allowed to forward to other relays, and a *time-to-live* value TTL that is utilized to remove obsolete messages. Furthermore, the header of unicast messages contains the destination user identifier, whereas, the header of interest-cast messages contains a threshold value α that is used to select the relevant destinations as explained in Section IV-B.

In PSNs nodes can exchange information as a communication opportunity arises. Accordingly, SANE procedures are triggered each time a node (say A) enters within the radio coverage of another node (say B). Initially, nodes exchange their interest profile (IP) as they will be used to take the most appropriate forwarding decisions, then each node start scanning its buffer of the messages to relay. The treatment of each message depends on its type, (i.e., unicast or interest-cast), and will be described respectively in Sections IV-A and IV-B. After all messages in the buffer have been analyzed, the node updates the buffer. This is achieved by

- *removing messages that are obsolete*: To this aim a deadline instant, t_{dead}, is assigned to each message in the buffer.
- *handling the messages relayed by the other node*: More specifically, if the node is a destination then the message will be forwarded to the application; if the node is a relay then it will insert the message in the buffer. As described above, a deadline instant t_{dead} is assigned to the message which is calculated as the value of the current time plus the TTL value reported in the message header.

A. Unicast

In the unicast case we aim at the best tradeoff between communication overhead and the probability of delivery success (i.e., the probability that the packet reaches the destination before it elapses), as well as the delivery delay. According to our interest-based approach, a message M should preferably be forwarded to individuals whose interest profile closely resembles the one of the destination.

More specifically, as in [21], we assume that in order to keep the communication overhead under control, the same message can be relayed at most for N^*_{replicas} times. Message relaying obey the following rules: Message M should be relayed to a node B if an only if both the two following conditions hold:

- the current value of N_{replicas} is higher than 1.
- the cosine similarity metric between the relevance of message M, denoted as $R(M)$, and the IP of B, denoted $IP(B)$, is higher than a given threshold ρ that we call *relaying threshold*, that is

$$\Theta(R(M), IP(B)) \geq \rho \tag{1}$$

The values of N_{replicas} and TTL contained in the message header are updated as follows: The value of N_{replicas} is halved, whereas the value of TTL is set equal to the difference between the deadline instant and the current time. Then, a copy of the message is sent to B. Note that, since N_{replicas} is equal to half the initial number of replicas at the sender node A, this is equivalent to handling node B half of the copies of M currently in node A's buffer, as done in BinarySW [21].

Obviously the message is transmitted to node B regardless of the value of N_{replicas} if B the destination of the message. In this case, node A will remove the message from the buffer after this is relayed to B.

The source is responsible of initializing the values of N_{replicas}, which must be a power of 2 and represents the maximum values of replicas of the message in the network, and the value of TTL, which represents the maximum delay acceptable for the delivery of the message. The message relevance profile is set equal to the interest profile of the destination.

Note that, as the threshold ρ decreases, the forwarding strategy becomes more aggressive. This results in the decrease of the delivery delay, and an increase of both the delivery success probability and the communication overhead (cost) incurred for the delivery of the message M, that we denote as $c(M)$. Observe that the cost $c(M)$ is proportional to the number of copies of the message M spread in the network. Note that a few extreme cases can be considered:

- $N^*_{\text{replicas}} = \infty$: in this case there is no bound on the number of copies of the message circulating in the network. We call the resulting version of our protocol suite *epidemic* SANE, and we denote it with SANE EP. The SANE version corresponding to the case $N^*_{\text{replicas}} < \infty$ is instead called *spray & wait* SANE and denoted SANE SW.
- $\rho = 0$: in this case, the relay threshold is not used, and the proposed forwarding strategy becomes the same as BinarySW [21]. Furthermore, if N^*_{replicas} is set equal to ∞ then our protocol behaves like epidemic forwarding [22], which is the policy achieving the lowest delivery delay (but also the highest cost).
- $\rho = 1$: in this case, only direct message delivery from source to destination is possible: Message delivery cost is minimized, but message delivery delay is very high.

In Section V, we will show the impact of the threshold ρ on the performance of the forwarding strategy through numerical examples.

B. Interest-cast

PSNs can create innovative services realized within the PSN itself, without the need of resorting to pre-existing communication facilities. *Interest-cast* is an example of such services in which a user wants to communicate a certain information (for instance, a movie at a local theater about opera composer Puccini) to the maximum possible number of interested users, within a certain time (e.g., the time of the last movie show). Interested users might have an interest in opera, or cinema, or both, and should be located in the "neighborhood" of the theater, so to be able to reach the theater if interested. This type of communication paradigm matches very well with the localized nature of PSN communications: the information is spread relatively fast in the neighborhood of the sender, while it takes longer to propagate to remote areas (which are typically less interested in the information, though).

Assume individual C wants to send a message M to all or the largest possible number of potentially interested individuals within the network. First, C must set the message relevance profile of M, which can be done assigning for each of the m interest dimensions a "relevance" value in the $[0, 1]$ interval. Such m-dimensional vector associated with a message is used (coupled with the individuals' interest profiles) to drive information propagation within the PSN. Note that the notion of message relevance profile allows to represent message M–similarly to individuals–as a point in the interest space. In the following, the relevance profile of message M is denoted $R(M)$. The set of *relevant destinations* for M, denoted $RD(M)$, is the set of individuals within the PSN for which message M is deemed relevant. As a consequence, $RD(M)$ is the set of nodes to which message M should be delivered, subject to an upper bound on the delivery time that we have called TTL^*. Whether a message M is relevant for a certain individual B is determined using a certain *relevance metric*. As we already explained, in this paper we use the well-know cosine similarity metric [6] to determine whether message M is relevant for individual B.

Note that, since both individuals' interests and message relevance profiles take values in the same m-dimensional interest space, we have that, for any individual B and message M, the angle between $IP(B)$ and $R(M)$ is in $[0, \pi/2]$, implying that $\Theta(B, M)$ is indeed in $[0, 1]$. In this paper, we use the following simple rule to determine whether message M is relevant to individual B: The message is relevant if and only if $\Theta(IP(B), R(M)) \geq \alpha$, where α is a suitably chosen *relevance threshold*.

We want to stress the difference between the notion of interest-casting defined herein and more traditional communication paradigms and services such as multi-casting and publish-subscribe. In interest-casting, the only action taken by a "content provider" (an individual generating a message) is determining the message relevance profile. After that, the message is injected in the network, and information propagation is driven by the notions of relevance and interest profile. As we shall see, these notions are used not only to dynamically determine the set of relevant destinations, but also to govern the forwarding process. Thus, in interest-casting the content-provider is not aware of the set of destinations the content should be delivered to, which is in sharp contrast with the traditional notion of multi-casting in which multicast groups are explicitly defined and typically known to the content provider. Furthermore, in interest-cast destinations must not explicitly subscribe to a specific "topic", as an individual is able to dynamically "capture" all (or most) relevant messages circulating in the PSN. This is also in sharp contrast with publish-subscribe mechanisms, which typically requires explicit subscription to one or more "topics" to be able to receive relevant information.

The forwarding discipline of interest-cast is similar in philosophy to the unicast case. In fact, if the two conditions given in Section IV-A for the unicast case hold then the message is relayed to B in the same way. If the above two conditions are not met but B is a relevant destination, then the message is transmitted with N_{replicas} set to one and TTL evaluated as explained in Section IV-A. Note that the above transmission does not have impact on the communication overhead.

V. Experimental Setup and Results

Here we present experimental results on the performance of SANE and UN-SANE compared to that of well known opportunistic forwarding protocols. For the evaluation we use both real-world traces (Infocom 06) and synthetic ones obtained with the SWIM mobility model [18]. We use also synthetic mobility traces to evaluate protocol performance because of the limited real-world traces enriched with user profiles, which does not allow evaluating performance under different conditions for what concern, e.g., the degree of correlation between individual meeting rates and similarity of their profiles. Such different mobility scenarios can instead be easily realized in SWIM by properly tuning the model parameters.

A. SANE vs Infocom 06

To validate the protocols on the Infocom 06 trace we average the results of the following experiment, repeated 100 times: We generate a message with a uniform traffic pattern (source-destination chosen uniformly at random), and we set message's relevance profile to be equal to the destination's interest profile. Then, we let the message to be forwarded in the network according to the different forwarding schemes. As already discussed in Section III-B, the correlation between node interest profiles and their meeting frequencies is low (see first row of Table II) without filtering out short meetings; on the other hand, filtering out short meetings to increase correlation would considerably reduce the size of the data set, making simulation results scarcely significant. In view of this, we have decided to keep the user population as large as possible (61 users, with a 0.08 meeting frequency correlation); thus, reasonably low values for the relay and relevance thresholds ρ and α should be chosen ($\rho = .25$ and $\alpha = .45$ in our case) for both unicast and inter-cast.

1) Unicast: We compare the unicast version of SANE (UN-SANE) to well known stateless forwarding protocols such as BinarySW [21] and Epidemic [22], and to a state-of-the-art of social-aware forwarding protocol, BUBBLE [11]. In implementing BUBBLE, we took care of putting the protocol in the best possible conditions, i.e., complete knowledge of the social graph and of the local/global ranking metrics. We consider both the SW and the uncontrolled version of UN-SANE in our experiments, denoted UN-SANE SW and UN-SANE EP, respectively. Being the network considered of only 61 nodes, parameter N^*_{replicas} (number of message copies) of BinarySW and UN-SANE SW is set to 4. The experiments are repeated for various values of the TTL's, and in each case, we measure the *average delay* (average delivery time for successfully delivered messages), the *cost* (average number of message copies in the network per delivered message, computed only for successfully delivered messages), and success percentage. The results are presented in Figure 2.

As can be seen, both versions of UN-SANE provide significantly higher success percentage than that of competing protocols (excluding, of course, Epidemic); also, the delay provided by the two versions of UN-SANE is better than that of both BinarySW and BUBBLE. In a sense, the two versions of UN-SANE provide different routing performance/cost trade-offs, with the SW version providing reduced success percentage with respect to the EP version (around 60% instead of about 68%), but with a much lower cost (factor 4 reduction in cost with respect to UN-SANE). Note also that the cost of UN-SANE SW is about the same as that of BinarySW, and only slightly higher than that of BUBBLE.

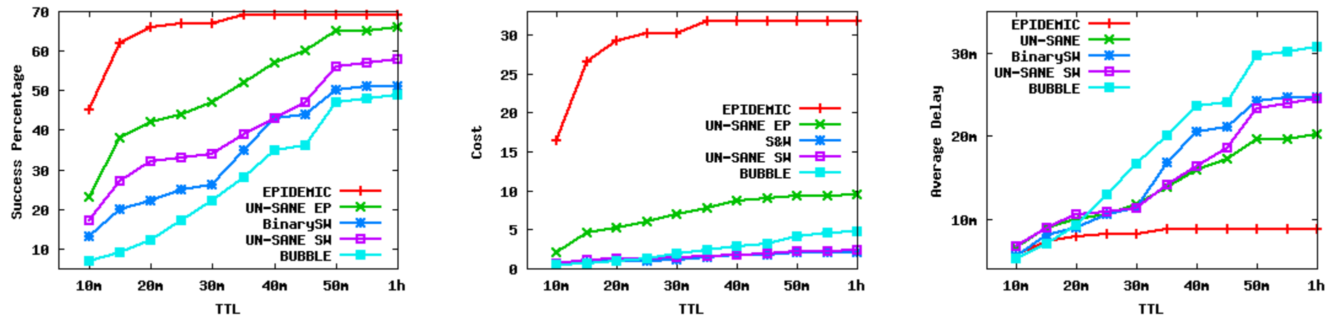
2) Interest-cast: Here, we show results related to the two interest-cast versions of our protocol: SANE SW, and SANE EP. Since there is no immediate way of extending BUBBLE into an interest-cast protocol, we compare SANE protocols only to Epidemic and BinarySW, whose interest-cast versions are straightforward (simply delivers a copy of the message to all relevant destinations). The way we generate messages and the input tuning parameters of BinarySW and SANE SW are the same as in the previous section. The results are shown in Figure 3. In this case, *coverage* refers to the percentage of relevant destinations holding a copy of the message when the TTL expires. As seen from the figures, SANE protocols perform very well, providing comparable coverage of relevant destinations to that of Epidemic (for TTLs values large than 30 min), but with a much reduced cost (as much as 10-fold cost reduction with respect to Epidemic, in case of SANE SW). The benefits of social-aware forwarding are evident comparing the relative performance of BinarySW and SANE SW: with a comparable cost, SANE SW provides higher coverage and lower delay as compared to BinarySW.

3) Limited Buffer Size: The previous results do not take in consideration possible limits on the size of node's buffer. Therefore, we have evaluated our protocols also with limited buffer size of the nodes, for different limits. Here, for the sake of space, we present only the results where the limit is 40 packets per node. Such results are shown in Figures 4 and 5. As you can notice, the performance and the delay of both UN-SANE and UN-SANE SW is not affected much by this buffer limit. On the contrary, the same limit badly affects the other protocols, and especially Epidemic (see Figures 4(a) and 4(c). Moreover, UN-SANE outperforms all the other protocols (including Epidemic) in terms of delivery success.

Also the multicast version of SANE reacts well to the limited buffer condition (see Figure 5). It outperforms all the other protocols in terms of coverage, except its SW alter-ego SANE SW for TTL's 25m–35m (see Figure 5(a)).

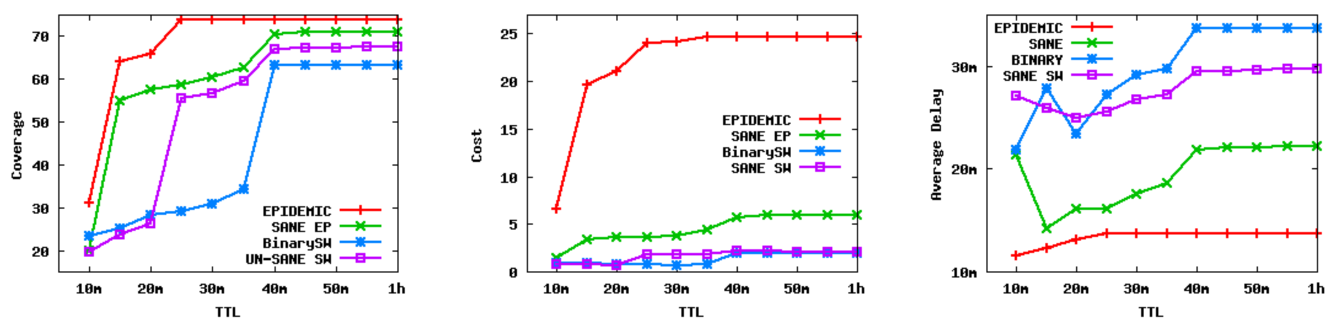
B. SANE vs Synthetic Traces

The synthetic traces we use for evaluation have been obtained from the SWIM mobility model [18]. In SWIM, nodes are assigned a home point in the network area, assumed to be a square. Each time a node has to choose its next destination,



(a) Success Percentage (b) Average Cost (c) Average Delay

Fig. 2. Performance of unicast protocols on Infocom 06 traces. Unlimited buffer.



(a) Coverage (b) Average Cost (c) Average Delay

Fig. 3. Performance of multicast protocols on Infocom 06 traces. Unlimited buffer.

it tradeoffs distance from its home point and popularity of the possible destinations. Thus, nodes with relatively close home points (*neighbors*) tend to go to the same locations and get in contact more often. For details on the SWIM mobility model, see [18], [19].

1) Experimental Setup: In order to run SANE on SWIM's traces, we do the following setup: First, we generate a the network, and a given number of network nodes. For each node, a 4-dimensional interest profile vector is randomly generated, with entries chosen independently and uniformly at random in $[0, 1]$. Each profile vector is then normalized to 1—this way, we make sure that no node has very low interests or no interests at all.

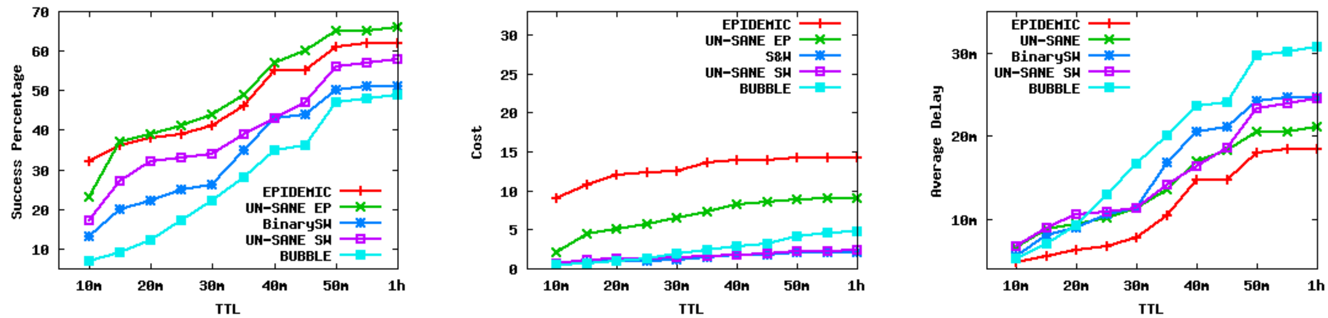
In SWIM, neighbors tend to have a higher meeting rate. The amount of correlation between vicinity of home points and meeting rate in SWIM is controlled by a parameter η: The higher this parameter, the higher this correlation will be. Thus, obtaining a relatively high meeting rate between nodes with similar profiles is easy: First we derive, for every node, its home point from the interest profile through a linear mapping, in such a way that nodes with similar profiles happen to be neighbors. This is done by using the first two coordinates of the profile as home-point coordinates. The correlation between profile similarity and home-point distances results very high (in our case it is -0.9). Then, we generate SWIM mobility traces, controlling the resulting correlation between node profile similarity and their meeting frequency by tuning SWIM's η parameter. Due to space limitation, in the following we will only show results for a SWIM simulation with $\eta = .9$, and 200 nodes scattered in a square area of $500m \times 500m$. The resulting correlation between interest profile similarity and pairwise meeting rates with these settings is about .7, allowing a wider range of variation for the relevance and relay threshold parameters of the SANE protocols.

Unfortunately, due to lack of space, here we do not present SWIM-based comparison results of SANE with the afore-mentioned well-known forwarding based protocols. Still we want to stress that due to the high correlation between node-profiles and pairwise meeting rates the advantage of the SANE protocols over the competitors becomes even more evident than in Infocom 06 simulations.

2) Varying SANE parameters: Once the value of the relevance threshold α has been set, the performance depends on the value of the relaying threshold ρ. In Figures 6 and 7 we show the success rate, average cost and average delay per received copy when the relevance threshold is $\alpha = .95$, versus the value of the relaying threshold ρ. As expected, the communication cost increases as the value of ρ decreases. This is obvious as a decrease of this parameter results in a less selective forwarding policy.

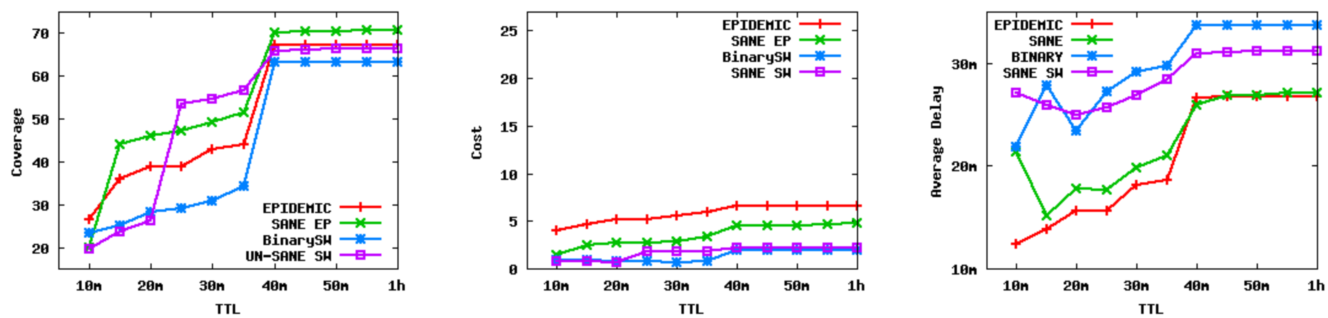
C. Discussion

When collectively considered, the experimental results presented in this section clearly show the superiority of SANE protocols over both social oblivious, stateless and social-aware, stateful approaches. Quite astonishingly, SANE provides better performance than competitors even when the degree of correlation between interest profile similarity and pairwise meeting rates is modest, as in the Infocom 06 scenario. If this correlation is higher, as it might be expected in practical



(a) Success Percentage (b) Average Cost (c) Average Delay

Fig. 4. Performance of unicast protocols on Infocom 06 traces. Buffer limit set to 40 messages.



(a) Success Percentage (b) Average Cost (c) Average Delay

Fig. 5. Performance of multicast protocols on Infocom 06 traces. Buffer limit set to 40 messages.

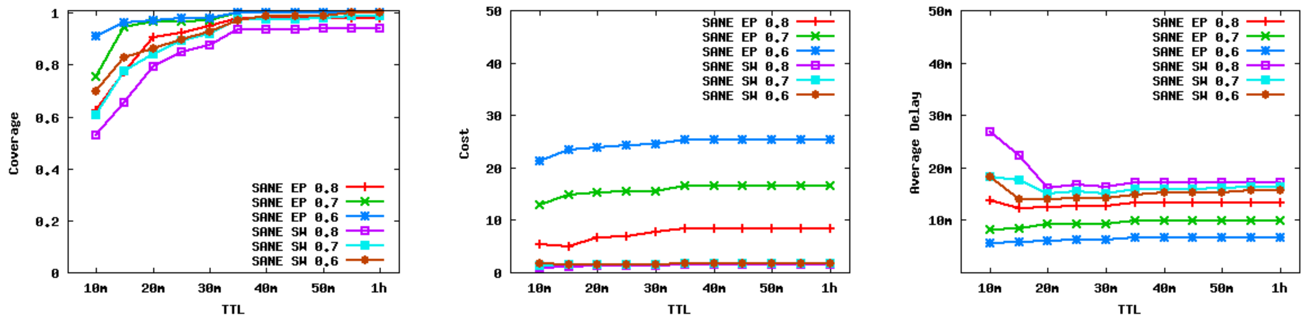
situations, advantages of SANE protocols over competitors become substantial. One might observe that indeed SANE protocols provide significantly higher average delay than Epidemic (although they have the lower delays as compared to the other competitors): however, it should be noticed that each packet is considered successfully received only if delivered with its TTL, a time which is deemed as acceptable for message delivery by the user sending the message. Thus, although the average delay of successfully delivered packets is higher with SANE as compared to Epidemic, this delay increase should be considered as acceptable by the users, since packets are still delivered to destination(s) within the TTL.

VI. Conclusions and Future Work

In this paper, we have first validated the intuition that individuals with similar interests tend to meet more often than individuals with diverse interests, and then used this intuition to design the first social-aware, stateless forwarding mechanism for opportunistic networks, called SANE. A nice feature of the SANE forwarding approach is that it can be used not only for traditional unicast communication, but also for realizing innovative networking services for PSNs, such as interest-casting. The results of extensive simulations based on both real-world and synthetic mobility traces have shown a clear superiority of our SANE approach over existing competitors. In particular, comparison with BinarySW clearly shows the benefits of social-aware forwarding.

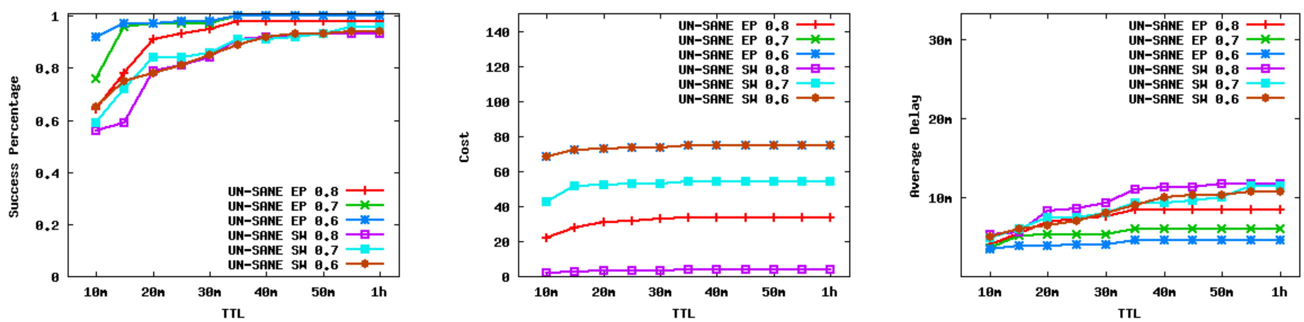
We believe the results presented in this paper open several avenues for further research. A first interesting research direction is trying to provide more extensive quantitative validations of the degree of correlation between interest similarity and pairwise meeting rates, which is likely to require suitably prepared mobility trace (and user profile) collection campaigns. Besides the research challenges, some practical issues need to be solved in order to actually realize SANE forwarding. Indeed, methodologies are required to identify the profile of users as well as the relevance of messages. User profile can be simply obtained if the user explicitly indicates which topics and to which extent she is interested. Such approach provides accurate user profiles bu is intrusive and the user might stop providing ratings unless she perceives a clear benefit. Alternatively, solutions can be applied that infer interests of users by observing how they browse information. For example, in [3] the relationship between user interest in a topic and the time spent reading a message regarding such topic is clearly demonstrated. Regarding message relevance, besides its explicit definition by the user generating the message other approaches can be used. For example, we may semantically analyze the message to be delivered (which may be inaccurate and adds a lot of complexity) or we may simply assume that users generate messages with relevance equal to their profile (which is extremely simple but inaccurate).

Finally, it is clear that the inferring and processing of user interest profiles poses a lot of privacy problems, just like all the other protocols that gather information on the social structure of the network. In the case of SANE, users share information in the process of computing their similarity. Fortunately, this can be done only sharing information about the intersection of their interests and it is thus possible to use the mechanism described in [15] that show how to do it in a private way. Nonetheless, privacy is still a largely open problem in pocket



(a) Coverage (b) Average Cost (c) Average Delay

Fig. 6. SANE in dependence of the relay threshold ρ.



(a) Succes Percentage (b) Average Cost (c) Average Delay

Fig. 7. UN-SANE in dependence of the relay threshold ρ.

switched networks.

References

[1] C. Boldrini, M. Conti, A. Passarella, "ContentPlace: Social-Aware Data Dissemination in Opportunistic Networks", *Proc. ACM MSWiM*, 2008.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms", *Proc. IEEE Infocom*, 2006.

[3] M. Claypool, D. Brown, P. Le, M. Waseda, "Inferring User Interest", *IEEE Internet Computing*, Vol. 5, No. 6, pp. 32–39. November/December 2001.

[4] P. Costa, C. Mascolo, M. Musolesi, G.P. Picco, "Socially-Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 26, n. 5, pp. 748–760, May 2008.

[5] E. Daly, M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs", *Proc. ACM MobiHoc*, 2007.

[6] M.M. Deza, E. Deza, *Encyclopedia of Distances*, Springer, Berlin, 2009.

[7] N. Eagle, A. Pentland, "Reality Mining: Sensing Complex Social Systems", *Personal and Ubiquitous Computing*, Vol. 10, n. 4, pp. 255–268, 2006.

[8] W. Gao, Q. Li, B. Zhao, G. Cao, "Multicasting in Delay Tolerant Networks: A Social Network Perspective", *Proc. ACM MobiHoc*, 2009.

[9] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks", *IEEE/ACM Transactions on Networking*, Vol 10, No 4, pp. 477–486, August 2002.

[10] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network", *Proc. of MobiCom*, 2004.

[11] P. Hui, J. Crowcroft, E. Yoneki, "BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks", *Proc. ACM MobiHoc*, 2008.

[12] P. Hui, E. Yoneki, S-Y. Chan, J. Crowcroft, "Distributed Community Detection in Delay Tolerant Networks", *Proc. ACM MobiArch*, 2007.

[13] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, "Pocket-Switched Networks and Human Mobility in Conference Environments", *Proc. ACM WDTN*, 2005.

[14] S. Ioannidis, A. Chaintreau, L. Massoulie, "Optimal and Scalable Distribution of Content Updates over a Mobile Social Networks", *Proc. IEEE Infocom*, 2009.

[15] L. Kissner and D. Song, "Privacy Preserving Set Operations", CRYPTO, 2005.

[16] F. Li, J. Wu, "LocalCom: A Community-Based Epidemic Forwarding Scheme in Disruption-tolerant Networks", *Proc. IEEE Secon*, 2009.

[17] M. McPherson, "Birds of a feather: Homophily in Social Networks", *Annual Review of Sociology*, vol. 27, n. 1, pp. 415–444, 2001.

[18] A. Mei, J. Stefa, "SWIM: A Simple Model to Generate Small Mobile Worlds", *Proc. IEEE Infocom*, 2009.

[19] S. Kosta, A. Mei, J. Stefa, "Small World in Motion (SWIM): Modeling Communities in Ad-Hoc Mobile Networking", *Proc. IEEE SECON*, 2010.

[20] A. Noulas, M. Musolesi, M. Pontil, C. Mascolo, "Inferring Interests from Mobility and Social Interactions, *Proc. Workshop on Analyzing Networks and Learning with Graphs*, 2009.

[21] T. Spyropoulos, K. Psounis, C.S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case", *IEEE Trans. on Networking*, Vol. 16, n. 1, pp. 77–90, 2008.

[22] A. Vahdat, D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", *Tech. Rep. CS-200006*, Duke Univ., April 2000.

[23] W. Zhao, M. Ammar, E. Zegura, "Multicasting in Delay Tolerant Networks: Semantic Models and Routing Algorithms", *Proc. ACM WDTN*, 2005.